

11) Publication number:

0 079 739

A2

12

EUROPEAN PATENT APPLICATION

(21) Application number: 82305926.6

(2) Date of filing: 08.11.82

(5) Int. Ct.²: C 12 N 15/00 C 12 N 1/00, C 12 P 21/02 C 07 H 21/04, C 07 C 103/52 //C12R1/19, C12R1/865

30 Priority: 12.11.81 US 320632

Date of publication of application: 25.05.83 Bulletin 83/21

Designated Contracting States:
 BE CH DE FR GB IT LI NL SE

7) Applicant: THE UPJOHN COMPANY 301 Henrietta Street Kalamazoo, Michigan 49001(US)

(72) Inventor: Dugaiczyk, Achilles c/o The Upjohn Company 301 Henrietta Street Kalamazoo Michigan 49001(US)

(72) Representative: Perry, Robert Edward et al, GILL JENNINGS & EVERY 53-64 Chancery Lane London WC2A 1HN(GB)

(S) Albumin-based nucleotides, their replication and use, and plasmids for use therein.

The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

ED 0 079 739 A2

GJE 70/2056/02

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in

5 development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly α-fetoprotein, but the synthesis decreases drastically after birth. Recently,

10 Law et al determined the complete sequence of mouse α-fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been 15 reached from studies on the α-fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al,

J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum 20 mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA comprises 2,078 nucleotides, template. The sequence starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-serleu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino 35 acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the tripledomain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by 20 labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleotides, of which 38 represent the 5'-untranslated region, 54 identify a prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since prepeptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a prepeptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the propeptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence located near the polyadenylation site has been identified by Renoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the concensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the human albumin mRNA (Table 1).

TABLE 1

						•			
	(30)	(170)	50 ala GCA (260)	(350)	(440)	(330)	170 gIn CAA (620).	(710)	(300)
5	AGC	20 1ys AAA		89 1 Eu	110 pro CCA	140 try 1AT		200 cys TCT	230 q1u GAA
	age T	asn phe AAT TTC	phe TT	ala thr GCA ACT	asn AAC	leu TTA	169 cys TGC	1ys AAG	ala GCA
]은 CTC		ale CA	al a GCA	ass GAC		168 cys TCT	leu CTC	phe TTT
	phe TTT	GA CA	val thr GTA ACT	70 phe gly asp lys leu cys thr val TTT CGA GAC AAA TTA TGC ACA GTT	asp	124 met cys thr sla phe his asp asn glu qlu thr phe leu lys lys tyr ATG FGC ACT GCT TTT CAT GAC GAG ACA TTT TTG AAA AAA TAC	9.4 9.4	P 4	alu GAG
	CTT CTT	gla GA	val GTA	ACA	AA	15 ₹	thr ACA	gln CAG	ela GCT
10	-10 ser leu leu TCC CTT CTT		3.5 5.5	57 57 50 50 50	S P S	leu lys TTG AAA	a ge	₹ ¥	1ys
	ser TCC	leu aly TTG GCA	AAT	leu 11A	qin his CAA CAC	phe TTT	ala	318	ور در
	11e ATT	lys asp AAA GAT	val GTG	AAA	leu 11G	th ACA	a) GCT	ser TCT	phe
	phe TT	178 AA	leu val TTA GTG	asp 1ys GAC AAA	phe	eAG GAG	AA AA	ser TCG	arg AGA
	p r o lys trp val tlu phe lle AAG TGG GTA ACC TTT ATT			ફે ઇ	100 101 glu cys phe leu gln his lys GAA TGC TTC TTG CAA CAC AAA	ale GAA		ale GCT	gln CAG
15	val c	10 arg phe CGG TTT	40 qlu asp his val lys gaa gat cat gta aaa	70 phe gly TTT CGA	100 101 alu eys GAA TGC	thr ala phe his asp asn glu ACT GCT TTT CAT GAC AAT GAA	160 arg tyr AGG TAT	190 lys ale AAG GCT	220 3er AGC
•	trp 766	ala his CCT CAT	S TS	Je CTT	asn	asa GAC	. ×.	91. 666	leu CTG
	-18 Met lys ATG AAG	val ala his GTT GCT CAT	qlu asp his GAA GAT CAT	Acc	lys gin glu pro aly ara asn AAA CAA GAA CCT GGG AGA AAT	E ST	160 ala lys arg CCT AAA AGG	e e	ويو دي
•	-18 Met ATG	glu val GAG GTT	€ &	rts CAT	41	phe TTT	a E	asp	ala CCT
20		916	pro phe CCA TTT	Jes CTT	Pro	al a	by 110	500	va1 GTA
20	1666	ser AGT	619 CCA		€ %	thr	1e2	leu CTT	e le SS
) (CTI	his lys ser CAC AAG AGT	34 gln cys CAG TGT	1ys AAA	lys gin glu AAA CAA GAA	124 met cys ATG TGC	leu CTC	916 GA	100
	CAG	h1s CAC		asp GAC	1ys AAA	met ATG	9.4 GA	asp GAT	£ 50
	(2)	a) a	gla	62 cys TGT	# # CCA	asp val GAT GTG	970 CCG	leu CTC	phe lys ala trp TTC AAA GCA TGG
25	GCTTTTCTCTTCTGTCAACCCCACAGCCCTTTGGCACA	p r o -1 1 val phe arg arg asp GTG TTT CGT CGA GAT	lea CTT	asn	91 cys TGT	asp GAT	္ ဧရ	180 pro lys leu CCA AAG CTC	210 ala phe lys ala trp GCT ITC AAA GCA TGG
	rctgi	arg arg	30 tyr TAT	95 S¥3	90 cys TGC	120 glu val GAG GTT		180 pro CCA	210 ala GCT
	rcti	arg CGT	ala gin CCT CAG	ala GCT	asp CAC	glu GAG	tyr phe TAC TTT	leu leu CTG TTG	و کو م
	1116	p r o val phe GTG TTT		ser TCA	ala	0 Y	tyr TAC		₹8
	ຮ	4a	ghe TT	ote GAG	met. ATG	ACA A	pro CCT	177 973 1GC	91 y CGA
30			918 GCC	asp GAT	glu met GAA ATG	leu val arg p TTG GTG AGA (arg his pro AGA CAT CCT	177 asp lys ala ala cys GAT AAA GCT GCC TGC	afg TT
		-1 -6 tyr ser arg gly TAT TCC AGG GGT	11e ATT	ala GCT	91y 567	leu 17G	87.9 ACA	ala GCT	1ys AA
		ser TCC	150 17G		tyr	pro arg CCC CGA	AG A	\$ \$	g Ay
	•	tyr	val	53 167 167	thr	5 5	e le CCC	asp	leu CTC
35		ala CCT	21 ala leu val leu IIe ala phe GCC TTG GTG TTG ATT GCC TTT	AG th	glu thr	leu CTC	141 glu ile ala arg GAA ATT GCC AGA	ala Cc	201 818 ser leu gin lys phe gly GCC AGT CTC CAA AAA TTT GGA
33		ser TCC	21 ala GCC	17s	81 879 CGT	111 asn AAC	₹ 5 8 2 8	171 ala CCT	201 ala GCC

	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
_	260 lev CTT	290 11e ATT	320 ala Cr†	350 ala GCC	380 leu CCT	410 arg CGT	A40 his cat	&70 ser AGT	500 173 AAA	530 val GTT
5	asp GAC	289 cys 7GC	tyr TAT] - T	pro CCT	val GTT	1ys AAA		<u>ور</u> (زر	le cri
	arq ala AGG GCG	hts CAC	asn AAC	879 AGA	1ys AAA	leu TTA	438 cys TGT	5 Y		اء درک
	arg	ser TCC	1ys AAA	leu CTG	phe TTT	Jeu CTG	437 cys TGT	thr	tyr TAC	ACT T
	asp GAC	leu leu alu lys CTG TTG GAA AAA			n es	a] a	\$ ¥	8,¥	ACA th	ala CA
10	asp CAT	leu leu alu lys CTG TTG GAA AAA	316 asp val cys GAT GTT TGC	val leu leu GTG CTG CTG	asp GAT	asn AAT	AGC	a fu GAG	2 4Y	Tys AA
	ala C:T	Jec 776	asp	val GTG	phe	ele CAG	val qly GTG GGC	leu his TTG CAT	es S	1ys AG
	253 cys TGT	leu CTG	1ys AAG	vál GTC	cre		val GTG		val GTC	ITE TYS ATC AAG
	250 253 asp leu leu glu cys ala CAT CTG CTT GAA TGT CCT	PTO CCT	310 phe val glu ser TTT GTT GAA AGT	367 TCT	ala lys val phe GCC AAA GTG TTC	400 glu tyr lys GAG TAC AAA	430 ser arg asn leu qly lys val TCA AGA AAC CTA GGA AAA GTG	val GTG	95 6	516 cys thr leu ser glu lys glu arg gln lle lys lys gln thr TCC ACA CTT TCT GAG AAG GAG AGA CAA ATC AAG AAA CAA ACT
	250 Jeu leu CTG CTT	280 qlu lys GAA AAA	310 val glu CTT GAA	tyr TAC	ala CCC	tyr TAC	95 A	460 461 1eu cys TTA TGT	Jeu CTG	874 AGA
15	250 Jeu CTG	280 atu GA	310 val CTT	340 88p GAT	370 tyr TAT	65 59 95 59	\$30 1ec CTA	460 1eu 17A	490 ala GCT	520 41u GAG
	asp CAT	278 279 cys cys TCC TGT	a t t	ala arg arg his pro GCA AGA AGG CAT CCT	369 his alu cys CAT GAA TGC	£ 55 7 43	AAC	P. CAO	phe ser TTT TCA	Jys AG
	917		asp CAT	arg his	9 to 6	leu	arg AGA	asn	phe TTT	26
	his CAT	of S	ala GCT	arg AGG	CAT ST	alu aln CAG CAG	ser TCA	Jeu CTG	pro cys phe CCA TCC TTF	ser TCT
20	245 246 cys cys TGC TGC	lys leu lys AAA CTG AAG	ala CCT	879 AGA	pro CCT		glu val	ser val val leu asn TCC GTG GTC CTG AAC		leu CTT
20	245 glu cys GAA TGC	lys leu AAA CTG	1eu		ala ala asp GCT GCA GAT	phe TTT	934	دي 10	arg arg Acc ccA	thr ACA
	ag &		ser 1CA	glu tyr GAA TAT	818 GCA	glu leu GAG CTT	val GTA	5	arg AGG	514 11e cys ATA TGC
	thr	ser	pro CCT		ala GCT	alu GAG	leu CTT	leu CTA	AAC	asp 11e GAT ATA
	h is CAC	367 100	leu TTG	tyr TAT	ala occ	392 cys 1GT	thr.	tyr IAT	val GTG	asp
25	val GTC	11e ATC	asp GAC	leu TTG	361 cys TCT	390 gln asn CAA AAT	Pro	450 glu asp GAA GAC	leu TTG	818 GCA
	240 1ys AAA	270 3er 10G	300 ala cct	330 phe TTT	360 cys TCC		420 thr		480 ser TCC	510 h1s CAT
	thr	asp	pro CC1	gly met GGC AÏG	glu lys GAG AAG	ile lys ATC AAA	val ser GTG TCA	\$ 50 80 80	ag 88	phe TIC
	1 5 TT	gla	met ATG					948 0ys 1GT	ACA th	thr
	asp GAT	glu asn GAA AAT	glu	leu 11G	glu thr thr leu GAA ACC ACT CTA	leu TTA	£ X	နို့ ဗ	476 477 0ys cys 1GC 1GC	phe TTC
30	leu val thr TTA GTG ACA	glu GAA	asp	phe TTC	thr	asn AAT	570 000	met ATG	476 0ys 1GC	ACA ACA
	4.4 GTG	265 cys TGT	glu asn GAA AAT	asp val	thr	ole CAG	val	ACA	thr 1ys Acc AAA	ala glu GCT GAA
	leu TTA	11e ATC	glu	asp GAT	914	pro CCT	17s	178	th Acc	818 CCT
	ser lys leu val TCC AAG TTA GTG	265 lys tyr lle cys AAG TAT ATC TGT	val GTG	ala lys GCA AAG	tyr TAT	gle GAG	thr lys lys val pro gln ACC AAG AAA GTA CCC CAA	gju ala lys arg met GAA GCA AAA AGA ATG	arg val AGA GTC	phe asn TrT AAT
25		1ys AAG	g Ju GAA	321 glu ala lys asp GAG GCA AAG GAT	thr ACA	381 val glu glu pro gln asn leu GTG GAA GAG CCT CAG AAT TTA	, the SA	9.1e		501 glu phe asn ala glu thr phe CAG TIT AAT GCT GAA ACA TTC
35	231 val CTT	261 ala CCC	291 ala CCC	321 910 GAG	351 1ys AAG	381 val GTG	tyr TAC	#41 pro CCT	471 83P GAC	501 91u GAG

	6	=	_
	(1790)	567 570 cys phe ala glu glu gly lys lys leu val ala ala ser gln ala ala leu gly leu ter ter TCC TTT GCC CAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCT TTA GAC TTA TAA CATCACATTTAAAAG (1883)	2W2
	560 178 AAG	¥	. ₹
5		ter TAA	ICTTI
J	558 559 cys cys TGC TGC	ACATI	EX
	AAG	CATC/	:ATA/
	alu GAG	Tè r	YAA
	val	14 E	CTAA
10	phe TTT	41,000	CTGT
	ala	13 E	CACC
	ala GCT	عاء درد	CCAA
	phe TTC	ala GCT	AAAG
15	550 83p GAT	580 gla CA	CTGT
15	asp CAT	ACT	SGTTC
	met ATG	ور ور	TT.
	val GTT	a 1 a	TTCT
	ala GCT	val	CTTT
20	<u> </u>	CTT	ATCT
) lev	AA A	ATTC
		,	(CCT
	29	£ 5	XXX
	~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~	8 4	AGAT C
25	S&0 Chr A ACA	570 8 91u c CAG	V1GA
	ရ ရေ (၁)	e ⊢	₹
		s d F	Ž
	* ñ	567 r eys c tec	₽ ₹
30	8 S	2 Q	ter ter \TCAGAATAAG
	₹ 5 \$ \$	6 9 6 9	A ter
	~ 2 ~ 3	8 ₽ .(∧	:TACC
	និ ក ន ច	رة بة ور بي) Yecc
	531 glu leu val 1ya hla 1ya GAG CTC GTG AAA CAC AAG	561 ala asp asp lya glu thr GCT GAC GAT AAG GAG ACC	ter ter Catctcagcctaccatgagaataagaagaagatgaagatcaaaagcttattcatctgttttcttttggttgg
35	v 03	v a G	ũ

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAAAAAATGGAAAGAATCTAA..... 20AA (2078)

Following are examples which illustrate procedures, including the best mode, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

Example 1 Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and 10 Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 680-685. 15

Example 2 Cloning Procedures

30

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczyk, A. (1980) Gene 10. 53-61]. It was annealed to PstI-linearized pBR322 DNA [Rolivar, F., 20 Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczyk, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, 25 S., et al., Ibid.]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [^{32p}]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HR101 by standard procedures well known to those of ordinary skill in this The E. coli RR1 hosts were lysed and then centrifuged to 35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris·HCl, pH 8.0, 10 mM CaCl₂, 10 mM MgCl₂). The cells for transformation are

prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HB101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl₂. Bacteria are then concentrated to one-tenth of this volume in CaCl₂ and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Rethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) <u>J. Virol.</u> 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) <u>Biochemistry</u> 11, 1242-1250] gels.

Example 4 DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and γ [^{32p}]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

Example 5 Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable <u>E. coli</u> host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

HB101(pHA206) - NRRL B-12550

5

E. coli HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public. upon the grant of a patent. It should be understood that the availability of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.

E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEp6 is a well known and widely available yeast episomal plasmid.

It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEp6) - NRRL B-12093.

Example 6 Assembly of the Serum Albumin Gene

Assembling the pieces together is a straighforward task of restriction enzymology. There is only one MspI site in the overlapping
DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI
digestion of the two DNAs, followed by (ii) the use of ligase, an
enzyme that seals DNA fragments, will give the desired product.
Although two other undesired DNA species will also be obtained in the
course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA
species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

(a) Escherichia coli

- (b) Saccharomyces cerevisiae
- (a) The vector of choice is plasmid pRR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoRl DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one of the yeast plasmid vectors, e.g., YEp6, at the Eco Rl cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the beginning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been documented that such **Escherichia** coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, 30 T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466. Screening of Clones Producing Albumin Example 8

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an <u>in situlysed</u> microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T. Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

CLAIMS

- 1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.
- 5
 2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.
- 3. E. coli HB101 (pHA36) having the deposit accession number $_{10}\,$ NRRL B-12551.
 - 4. $\underline{\text{E. coli}}$ HB101 (pHA206) having the deposit accession number NRRL B-12550.
- 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

	(30)	(170)	(092)	(350)	(044)	(330)	(629)	(017)	(300)
	AGG	178 AA	818 GCA	8 3 E	110 CCA	140 try TAT	170 aln GAA	200 cys TCT	230. glu GAA (
5	ohe TTT	phe TTC	phe TIT	thr	AAC	leu TA	169 cys TGC		8 g g
5		asn	alu CAA	a1a CCA	330	tyr TAC	168 cys 1CT	leu lys CTC AAG	a be
	phe TTT	₽. 6¥	thr ACT		asb GAT		2 8		Ske n
	3 E	2 ₹	glu val GAA GTA	ACA th	1ys AAA	1ys AAA	th ACA	aln ara CAG AGA	ala GCT
	5	శ్రీ క్ర	₹.	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	ohe leu ain his lys TTC TTG CAA CAC AAA	phe leu lys lys TTT TTG AAA AAA	phe 111	8 %	
10		lys asp leu AAA GAT TTG	asn	7 E	ain his CAA CAC	phe I	ala GCT	ele lys GCC AAA	pro lys CCC AAA
		asp GAT	val	13.5 AA	lec 11G		ala CCT	ser TCT	phe TTT
		lys AA	1 € 1 €	88D GAC	phe TTC	950	lys AA	36. 100	870 AGA
		phe TT	40 val lys GTA AAA		10 10 10 10 10 10	5 ₹3	tyr TAT	ele GCT	220 leu ser gln arg CTG AGC CAG AGA
• •		10 his arg CAT CGG	40 his val cat gta	70 phe TTT	5 £ 5 X	130 asn AAT	160 819 AGC	190 173 AAG	220 ser gln AGC CAG
15		his CAT	7 TS	5 5	asn AAT	8.5 GAC	1 ys	91. 666	ned CTG
		#13 CCT	859 CAT	thr	2 2	h I s	ata CC	# 3	ala arg ccr ccc
		glu val GAG GTT	£ 8	h1s CAT	2 88	phe TTT	oke TIT	asp GAT	ala arg ccr ccc
		gju	ag r	leu CTT	glu pro GAA CCT	ele GCT	phe TTC	. 500 000	val GTA
20		Ser AGT	5 Y	ser TCA	g g	thr	leu CTT	le CTT	ala trp ala val GCA TGG GCA GTA
		his lys CAC AAG	34 673 TGT	asp lys GAC AAA	ele CAA	124 cys TGC	glu leu GAA CTC		t 70
		h I s	gla	asp GAC	1 ys	met ATG		asp GAT	ද
		ala GCA	gln CAG	62 cys TGT	818 CCA	val	5 5	le CTC	<u>}</u>
		asp GAT	CT CT	asn AAT	91 Cys	asp	. els	1ys AG	phe TTC
25		 829	30 tyr 1A1	8 2 3	85 160	120 val GTT	150 tyr TAT	85 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	210 ela GCT
		arg CGT	gln CAG	ala GCT	8 5 CAC	5 S	age 1	leu TTG	\$ 5
		p r o val phe CTC TTT	ala ccr	ser TCA	ala 601	5 Y	tyr TAC	Jeu CTG	916 S\$
		p r o gly val phe a CCT CTC TTT C	phe TTT	910	met ATG	3rg 405	679 CC 1	17. 673 700	
30		gly GGT	18 000	88P GAT	glu met GAA ATG	leu val TTG GTG	h is	2 20	Phe TTT
30		-6 arg ACC	11e ATT	ala GCT	91y CCT	leu TTG	P S	818 CCT 0	1ys AA
		-1- 367	15 176	53 oys val TGT GTT	tyr TAT	£ 55	\$ 5 \$	1 × ×	£ 8
		tyr	val GTG	53 038 1GT	thr	ະຄຸ	ala arg arg his pro CCC AGA AGA CAT CCT	ele esp GCT GAT	ser leu gin lys phe gly AGT CTC CAA AAA TIT GGA
		ala GCT		thr ACA	gle GA	leu CTC	II.e	1 50 50 50 50 50 50 50 50 50 50 50 50 50	ser AGT
35		36r 706	21 ala ccc	2 2 X	8 67 700	111 asn AAC	± 58	171 818 GCT	201 ala GCC

	(890)	(086)	(10701)	(1160)	(1250)	(1340)	(1630)	(1520)	(1610)	530 val GTT (1700)
5	260 1ec CTT	290 11e ATT	320 ala GCT	350 ala CCC	380 leu CCT	410 819 CGT	\$40 h1s CAT	\$70 Ser AGT	Sno 1ys AAA (530 val CTT (
	ala asp GCG GAC	289 his cys CAC TGC	tyr TAT	leu CTT	pro CCT	val GTT	1ys AAA	Val		530 leu val CTT GTT
		الم در	asn	37.8 AGA	phe 1ys TTT AAA	leu TIA	438 cys TGT	מנא ו	val p	ain thr ala leu CA ACT GCA CTT
	₽ L 0.00 0.00	107 107	316 oys lys TGC AAA	leu CTG	ag E	Jeu CTG	437 438 cys cys TGT TGT	thr 1	tyr v	2 t
	345	glu lys GAA AAA	316 val cys lys GTT TGC AAA	Jeu CTG	ale GAA	هاء 20		lys (ein thr CAA ACT
10	ala esp asp arq GCT GAT GAC AGG	pro leu leu glu lys ser CCT CTG TTG GAA AAA TCC	val	Jeu CTG	asp CAT	88n AAT	430 asn leu qly lys val qly ser lys AAC CTA GGA AAA GTG GGC AGC AAA	461 oys val leu his qiu lys thr pro TGT GTG TTG CAT GAG AAA ACG CCA	alu thr GAA ACA	\$.₹
		leu TTG	85D GAT	val GTG	phe	1 1 1 CAG	, yls	7. ST 0	asp q CAT G	86 A
	253 glu eys GAA TGT	leu CTG	1ys AAG	val GTC	val c	phe TTC	18 c	1eu 7		5 A
	3.8	280 glu lys pro leu leu GAA AAA CCT CTG TTG	Ser AGT	ser TCT	ala lys val phe GCC AAA GTG TTC	¥ %	1 × ×	val 1 GTG 1	glu vəl GAA GTC	514 818 839 11e cys thr leu ser qlu lys glu arq qln 11e lys lys GCA GAT ATA TCC ACA CTT TCT GAG AAG GAG AGA CAA ATC AAG AAA
15	250 leu leu CTG CTT	173 AA	g tu	tyr TAC	8 ts CC	\$00 gly glu tyr lys GGA GAG TAC AAA	1, V 1	461 0ys v TGT 0		5 3 5 3
15	250 1eu CTG	280 91u GAA	310 val cr7	340 asp GAT	370 tyr TAT	400 glu tyr GAG TAC	430 1eu c	460 4 leu o TTA T	490 ala leu GCT CTG	514 11e cys thr leu ser qlu lys glu arq ATA TGC ACA CTT TCT GAG AAG GAG AGA
	gly asp GGA GAT	279 6ys TCT	phe TTT	5 T	369 Cys	¥ 69	AC C	460 gln leu CAG TTA		A 26 50
		278 cys TGC	350	ala arg arg his GCA AGA AGG CAT	369 qlu cys GAA TGC		5 2		phe ser TTT TCA	25
	245 246 6ys 6ys his TGC TGC CAT	lys leu lys glu AAA CTG AAG GAA	ala GCT	arg AGG	h1s CAT	d Ju	val ser arg GTC TCA AGA	leu ser val val leu asn CTA TCC GTG GTC CTG AAC	cys p 1GC 1	5 d
20	246 673 760	lys AAG	ala GCT	arg	asp pro his GAT CCT CAT	o o o		ָ בֶּעָ זיר ט	pro c CCA T	3 =
	245 633 160	Jeu CTG	T Y	\$ T\$	asp pro GAT CCT	phe 111	glu val GAG GTC	val val GFG GFC	arg p	ر د ج
	545			tyr TAT	ala GCA	Jeu CTT	val GTA	, b	arg a	514 cys t TCC A
	his thr cac acc	Ser	leu pro TTG CCT	g &	ata GCT	glu leu CAC CTT	pro thr leu val CCA ACT CTT GTA	leu ser CTA TCC	asn arg AAC AGG	Z e Z
	h 18	36. 100	1eu 11G	tyr TAT	8 500	392 cys TGT	thr	tyr 1 TAT 0	val a	SP 1
25	val GTC	11e ATC	300 ala asp ccr cAc	Jeu TTG			Pro CCA	asp t	1eu v 11G G	. 25
	240 1 ys	270 ser TCG	300 ala CCT	330 phe TTT	360 375 155	390 gin asn CAA AAT	420 thr	450 91u a	480 ser leu val TCC TTG GTG	510 phe his ala asp TTC CAT GCA GAT
	leu thr CTT ACC	asp CAT	met pro	gly met GGC ATG	glu lys	AA	77	618 67A	glus GAA T	5 phe h TTC C
	leu CTT	gt CA	met ATG	91. SSC 5	glu GAG	11e	/al /	648 078 1GT G	£ 5	ر ب م ب
30	asp GAT	glu esn GAA AAT	glu	Jec TG	Jeu	TZ	45	តិ ភិ ភ	477 cys t TGC A	5 T €
50	th ACA	91° GA	asp GAT	af T	thr Act	esn AAT ;	. E S	16 P	476 4 0ys 0 TGC T	ر ا ا
	val GTG	265 cy3 TGT	asn AAT	va1 GTC	tyr glu thr thr leu TAT GAA ACC ACT CTA	gin asn leu ile CAG AAT 7TA ATC	15 T	₽5 -	¥3 ¥0 ¥	25
	led 17A		88	asp	38	glu pro GAG CCT 0	\$.₹	\$ % \$ %	7 S	5 G
	1ys Ag	tyr lle tat atc	V&1 CTC	lys AG	tyr TAT	2 S	ys 1 AG /	- Y	2 C	
35	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	261 818 lys GCC AAG	giu val giu asn asp giu GAA GTG GAA AAT GAT GAG	a1a	thr ACA 1	914 g	- S	448 glu ala lys arg met pro oys GAA GCA AAA AGA ATG CCC TGT	476 477 arg val thr lys oys cys thr AGA GTC ACC AAA TGC TGC ACA	7 92 17 A2
	231 val GTT	261 818 CCC	291 818 GCC	321 glu ala lys asp val phe leu GAG GCA AAG GAT GTC TTG TTG	351 Iys thr AAG ACA	381 val 9 GTG 0	411 tyr thr lys lys val pro gln val aer TAC ACC AAG AAA GTA CCC CAA GTG TCA	441 pro g CCT G	471 88P BI	501 glu phe asn ala glu thr phe thr GAG TIT AAT GCT GAA ACA TIC ACC

	6	33	(2)
	(1790)	188	120
	560 173 AAG	Ծ	¥
5	8 8 5	ter	717
)	558 559 cys cys TCC TCC	CAT	ATT
•	Lys .	ATC	ATA
	550 1ys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys cys c AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT CCT TTT GTA GAG AAG TGC 1	ala leu qly leu ter GCC TTA GGC TTA TAA CATCACATTTAAAAG (1883)	AAAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTCTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAACATAAATTTFTTTAA (2002)
	CT A	T Y	TAA
10	& E	÷ 8	757.
	1 50 E	3 5) YCC
	sls GCT	- 2 - 2 - 2 - 3	CAAC
	T e	ala CCT	¥ ¥¢¢
	550 • 55 CAT	580 91n	GTGT,
15	8 5 CA T	AGT	GTTG
	met ATG	818 GCA	1110
	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	ala	TCTT
	# CC	va] GTT	1111
20	\$ ₹	Jeu CTT	10.00
	leu	1ys AAA	17CA
	es es	lys AAA	CTTA
	95 86 86	a) y GGT	¥¥
	\$ \$	9 6 86	ATCA
25	S&O thr	570 phe ala glu glu gly lys lys leu val ala ala ser gln ala ala leu gly leu ter TTT GCC GAG GAG AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA	GA6
	# Y20	253 SCC	WA7
	1ys AG	P F T T T	₹
	. 25	567 cys TGC	ชั
20	1ye AAG	thr Acc	ter ter
30	• 1 0	glu GAG	5
	leu val lys his lys pro	567 85p 85p lys glu thr cys GAC GAT AAG GAG ACC TGC	ζ Ž Ž
	4	asp TXT	CTAC
	leu CTC	35 C	<u>ک</u> کی
35	531 9lu GAG	sela ala d CCT C	ter ter CATCTCACCCTACCATGAGATAAGAGA

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAAAAAATGGAAAGAATCTAA..... 20AA (2078)

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

5	(170)	(260)	(350)	(440)	(330)	(620)	(710)	(300)
	20 1ys AAA	50 818 67A	90 Jeu CTT	110 CCA	140 try TAT	170 41 CAA	200 cys TGT	230 glu GAA
	asn phe AAT TTC	val thr glu phe GTA ACT CAA TTT	ala thr GCA ACT	AAC	1 te	169 cys 1GC	1ys AG	818 GCA
	AAT	alu CAA		35 CAC	tyr TAC	168 cys TGT		
10	alu alu GAA GAA	thr	thr val	asp CAT	1ys	al Ay	2 VOV	alu phe GAG TTT
	1 P. S.	glu val GAA GTA	th ACA	1ys AAA	 \$.₹	ACA th	CAG 93	ala CCT
	aso leu aly GAT TTG GGA	2. 4S	75 asp lys leu cys GAC AAA TTA TGC	bhe leu gin his lys asb TTC TTG CAA CAC AAA GAT	phe leu lys TTT TTG AAA	a be	1ys	phe pro lys ala TTT CCC AAA GCT
	leu TTG	AAT	15 T	ala CAA	phe TTT	818 GCT		200
		val GTG	1ys	1ec	thr ACA	sla sla GCT GCT	ser ala TCT GCC	phe pro TTT CCC
15	173 AA	leu TTA	GAC		56		367	
	10 ala hís arg phe lys GCT CAT CGG TTT AAA	40 val lys leu val GTA AAA TTA GTG	45 68	100 101 91u cys GAA TGC	130 asn glu glu thr phe AAT GAA GAG ACA TTT	160 phe ohe ala lys arg tyr lys TTC ITT GCT AAA AGG TAT AAA	ala GCT	pln arg CAG AGA
	10 01 000	40 val GTA	phe TTT	100 F 4 5	130 asn glu AAT GAA	160 AGC	190 173 AAG	220 3er AGC
	ala hís GCT CAT	MIS CAT	thr leu ACC CTT	gly arg asn	asp GAC	AAA	190 alu aly lys GAA GGG AAG	Jeu CTG
20	ala GCT	asp CAT		1 5	h is	et e	45 64	ala arg leu GCT CGC CTG
20	his lys ser glu val CAC AAG AGT GAG GTT	2 ₹	leu his CTT CAT	9) y 000	ala phe GCT TTT	ohe	asp GAT	ala GCT
	26.2	pro phe CCA TTT	Jeu CTT	pro CCT		phe TTC	arg 000	
	AGT	SCA CCA	36. TCA	916	124 cys thr TGC ACT	leu leu CTC CTT	leu CTT	
	17. AG	34 gln gln cys r CAG CAG TGT	62 cys esp 1ys TGT GAC AAA	gin	124 cys 160	lec CTC	gf.	trp ala TGC GCA
25	his cac	g J	889 GAC	ala lys GCA AAA	val met GTG ATG	ole GAA	asp GAT	
	1 88P 8la GAT GCA	g],	62 9ys TGT		val GTG	500	S C C C	lys ala AAA GCA
	esp GAT	30 gln tyr leu CAG TAT CTT	AAT	90 91 cys ays TGC TGT	asp	္ ဦ	1ys AAG	Phe
		30 tyr TAT	60 91u GAA		120 val GTT	150 171	180 CCA	210 818 GCT
			918 GCT	asp GAC	5 S	tyr phe TAC TTT	1eu 776	
30		els GCT	16. 10.	ala CCT	5 50 CCA 50	1 XC	CTG CTG	glu arg GAA AGA
		phe TTT	gle SAG	glu met CAA ATG				915
		e je 000	asp GAT	glu met CAA ATG	val arg GTG AGA	5 E	#1# 200	
		11e ATT	val ala GTT GCT	aly ccr	Jeu 776	\$ 54 \$ 54	177 ala ala oys GCT GCC TGC	AA A
35		leu TTG	va1 GTT	tyr	r S	87.9 AG AG	1 × 3	g]n CAA
		21 ala leu val leu lle ala phe GCC TTG GTG TTG ATT GCC TTT	51 53 lys thr cys val ala AAA ACA TGT GTT GCT	glu thr GAA ACC	leu pro arg leu val arg CTC CCC CGA TTG GTG AGA	ile ala arg arg his pro ATT GCC AGA AGA CAT CCT	85P lys GAT AAA	201 81a ser leù gln lys phe GCC AGT CTC CAA AAA TTF
	•	Jeu 776	th ACA	£8	leu CTC	ATT	ala CCT	AGT
		21 818 CCC	25 TX	81 879 CGT	111 88n AAC	14 g 8	171 818 CCT	201 818 GCC /

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 leu CTT	290 11e ATT	320 818 CCT	350 ala GCC	380 leu CCT	\$10 \$70 CGT (480 his cat (470 3er AGT (
	83P CAC	289 cys	tyr	7 E	Pro 1		Iya h			530 u val
	818 600	77 S	AAC	5 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	Jys p	leu val TTA GTT	438 cys 1; TCT A/	o val A GTA	1 pro	a leu CTT
10	5. S	10.	1ys	leu a	phe 1 TTT A	Jeu Je	437 43 cys c) TGT TC	r 870 G CCA	r val	thr ala leu ACT GCA CTT
10	320	2.3		leu l	glu p GAA T	818 J	lys cy AAA TC	s thr A ACG	1 ty 1 AC	
	asp asp ang ala GAT CAC AGG GCG	glu lys ser GA AA TCC	316 val cys GTT TGC	leu leu CTG CTG	asp q GAT G	asn al	ser lys	u lys	alu thr GAA ACA	£ 3
	ala CCT	3 5	316 ser lys asb val cys AGT AAG GAT GTT TGC	val 1 GTG C	phe asp TTC GAT	gln asn CAG AAT	aly ser GGC AGC	leu his glu lys TTG CAT GAG AAA		ain ile iys iys ain CAA ATC AAG AAA CAA
	253 267 167	pro leu leu CCT CTG 1TG	lys A					u his	SAT CAT	¥ 3
15	S ale	2 5	ser 1 AGT A	ser val TCT GTC	Tys va	lys phe AAA TTC	lys val AAA GTG	1 leu G 17G	glu val GAA GTC	11e
	CTT	lys pro leu leu AAA CCT CTG TTG	glu s	tyr s TAC T	ala lys val GCC AAA GTG			1 s val T GTG		5 5
	250 leu leu CTG CTT	280 q1u 1 CAA A	310 val g CTT C	340 asp to	370 tyr al	400 glu tyr GAG TAC	0 4 93 4 664	460 461 leu cys TTA TGT	490 ala leu GCT CTG	520 glu arg GAG AGA
	asp 1	279 2 cys q TGT Q	Dhe v	Pro a	369 3; cys ty TGC 17		430 n 1eu		818 GCT	520 g1u GAG
	93, 8		asp p	his pi CAT CO	alu ey CAA TC	u 93y T GGA	arg asn AGA AAC	gin CAG	3er TCA	glu lys CAG AAG
20	hls g	910 G	ala a	arg h	his al	aln leu CAG CTT	7 A	J asn	phe TTT	916
	246 cys h TGC 0	1ys g	ala a GCT O	arg a	369 pro his alu cys CCT CAT GAA TGC	u aln G CAG	I ser C TCA	1 leu : CTG	633 1GC	ser TCT
	245 678 160 1	278 Leu lys glu cys CTG AG GA TCC	160 174 0	ele arg GCA AGA	asp pr	e glu	glu val GAG GTC	ser val val leu asn TCC GTG GTC CTG AAC	5 A	514 cys thr leu ser glu lys TGC ACA CTT TCT GAG AAG
	SA 16 A				25 este	u phe	ole 1	val GTG	5	thr
05	thr g	ser lys AGT AAA	pro ser CCT TCA	glu tyr GAA TAT	a ala T GCA	glu leu GAG CTT	, val		arg AGG	514 Cy3
25		ser s	leu pi		a ala		thr leu val ACT CTT GTA	tyr leu TAT CTA	asn AC	11e ATA
	val his GTC CAC	11e s ATC T	asp 1c CAC TI	leu tyr TTG TAT	s ala T GCC	392 n cys r TCT			val GTG	esp
	240 1ys v AAA G	270 ser 11		0 e 1eu T 77G	0 361 s cys c 161	o esn	420 thr pro ACT CCA	asp GAC	Jec 116	818 GCA
*	thr 13	23 839 SC		330 t phe c TTT	360 3 cys 2 TCC	390 1 gln		97u 97u GAA	\$80 567	510 h1s CAT
30	lea ti	# 55 E X3	t pro	y met C ATG	glu lys GAG AAG	¥	ser TCA	عاء 20	glu GA	phe TTC
_			u met G ATG	91y 5 550		11e	v31 GTG	448 073 1GT	th	thr
	# 50 12 € 55	glu asn GAA AAT	e gle CAG	phe leu TTC TTG	r leu r cta	3 E .	gla CAA	5 22	477 cys 100	phe TTC
	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	2	n 889 T CAT	á.E	glu thr thr leu GAA ACC ACT CTA	glu glu pro gln asn leu GAA GAG CCT CAG AAT TTA	. 579 573	glu ala iya arg met GAA GCA AAA AGA ATG	476 477 1ys cys cys AAA TGC TGC	glu thr phe GAA ACA TTC
	3 €	265 e cys C TGT	u asn A AAT	val GCC	thr Acc	18 A	val	arg ACA	<u>₹</u>	gle CA
35	ser lys leu TCC AAG TTA	r 11e T ATC	1 91v 6 GAA	dsp CAT		glu pro GAG CCT	lys lys AAG AAA	ala lys GCA AAA	thr	ala GCT
		s tyr	glu val GAA GTG	ele lys CCA AAG	tyr TAT	gt. GA	1ys AAG	818 GCA	val	asn
		261 ala lys GCC AAG		321 glu ala lys GAG GCA AAG	JS1 lys thr tyr AAG ACA TAT	n16 ₩	411 tyr thr lys lys val TAC ACC AAG AAA GTA	ole GA	۲ کو کو ع	501 glu phe asn ala GAG TTT AAT GCT
	231 val GTT	261 ala GCC	291 a18 GCC	321 91 u GAG	351 1ys AAG	. 381 val GTG	411 tyr TAC	pro CCT	asp GAC	91c

5	(1790)	570 phe ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter TIT GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG (1883)	JAAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTCTTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAACATAAATTTCTTTAA (2002)
	560 178 AAG	, AG	¥ .
	559 560 cys 1ys TGC AAG	ter	TCTI
	558 100 100	ACAT	AATT
10	1ys AAG	CATG	CATA
10	914	ter TAA	AAAA
	558 550 1 550 55	Jeu TTA	ICTAA
	ag II	5 5	CTG
	ala GCT	15 T	CACC
15	ete GCT	# T# 000	χχ
	phe TTC	ela GCT	.WYG
	350 385 CAT	580 91n CAA	SCTG
	asp GAT	Ser	CCTT
	met. ATG	818 GCA	Ĕ
20	val GTT	318 GCT	1101
	ele GCT	val GTT	GTTT
	% 8	ieu CTT	ATCT
	CTC	% ¥ *	ATTC
25	e de CAA	₹ 8	2011
	91c GAG	91y GGT	¥
	£.¥3	95	CATC.
	\$ t \$ 6	578 916 6AG	TGAA
	8 t 8	a 1 CC	¥
30		phe TTT	AAAG
	21 0	567 cys TGC	766
	lys AAG	thr	ter ter ATGAGAATAAC
	h1s CAC	olu GAG	ter
35	531 glu leu val lya hia lya pro GAG CTC GTG AAA CAC AGG CCC	567 5 asp lys glu thr cys (C GAT AAG GAG ACC TGC	ter ter Catctcagcctaccatgagaataagaga
၁၁	val	asp CAT	CCCT
•	leu CTC	£ 3	CTCA
	55 55 56	561 ala GCT	2

TCATTITGCCTCTTTTCTCTGTGCTTCAATTAAAAATGGAAAGAATCTAA.... 20AA (2078)

Nucleotide sequence coding for the prepeptide of human serum 7. nce is as follows:

	albumin,	said	nucleotide	sequenc
				(30)
5				ser AGC
				ohe TTT
				phe lev TTT CTC
				a phe
10	,			ser leu leu p TCC CTT CTT T
				1 in 12 in 1
				1 s
				phe 11e
15	•			110 p
				p r o val tlu g GTA ACC 1
				tr.
				1 ys
20				-18 Met
20				SCACA
				110
				رورزو
				נכעכי
25				YACC
				TGTC
				-18 Het GCTTTCTCTCTGTCAACCCCACAGCCCTTTGGCACA ATG
	•			TTTC
30				g

ser ala tyr ser arg gly val phe arg ard TCG GCT TAT TCC AGG GGT GTG TTT CGT CGA

35

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

					•			
5	(170)	(192)	(350)	(440)	(330)	(620)	(017)	(300)
	20 173 AAA	S = 5	80 Jeu	110 pro CCA	140 try 1AT	170 91n CAA	200 cys	230 91u 6A
	phe TTC	phe TTT	thr	AAC	1eu 17	169 cys TGC		818 GCA
	AAT	5 AA	618 GCA	330	TAC	168 0ys TGT	leu lys CTC AAG	phe TTT
10	g in	thr	val GTT	asp CAT	1ys AAA	2€	P 20	atu
10	4 S	glu vəl GAA GTA	th T	Tys AA	1ys AA	thr	aln ara CAG AGA	ala CCT
	leu aly TTG GGA		75 leu oys TTA TGC	his CAC	leu TG	. ag E	1 7 3	1ys
	1ec 176	AAT	leu TTA	CAA 91	phe TTT	ata GCT	ala lys GCC AAA	970
	SAT	leu val TTA GTG	AAA	leu TTG	A th	a)a GCT	ser TCT	ohe TTT
15	phe lys TTT AAA		88 CAC	phe TTC	5 9	17s	ser TCG	arg AGA
	age T	40 val lys GTA AAA	91 <i>y</i> CGA	101 8%9 10C	g to			gla
	10 819 CGG	40 CTA	24 phe	100 GA 6	130 asn AAT	160 arg tyr AGG TAT	190 lys ele AAG GCT	220 AGC AGC
	glu val ala his GAG GTT GCT CAT	55	thr leu ACC CTT	asn AAT	asp GAC	lys AA	91. 555	leu CTG
	sls GCT	asp GAT	leu his thr CTT CAT ACC	arg AGA	h is	8 ts	age 45	وته رون
20	glu val GAG GTT	olu GAA	ser leu his TCA CTT CAT	91y 666	phe 177	1 te	asp	ela GCT
	91 ⁶	phe TTT	leu CTT	pro	818 GCT	leu phé CTT TTC	و د د د	val GTA
	lys ser AAG AGT	Pro		£ 8	thr ACT	glu leu leu phé bhe GAA CTC CTT TTC TTT	leu CTT	ala SCA
		34 gln ays CAG TGT	asp lys GAC AAA	£87.	124 cys TGC	leu CTC	91°	100
25	ele his GCA CAC			lys AA	aet ATG	g ^g g.	lys leu asp AAG CTC GAT	# 55 SC A
		gln CAG	62 935 1GT	్లో స్ట్ర	val GTG	2 000	c c	1ys
	1 esp CAT	leu CTT	asn	91 cys TGT	asp	1 6	1ys AAG	phe TTC
	 500	30 tyr TAT	69 16 AA	98 973 167	120 val GTT	150 tyr TAT	180 pro	210 ala GCT
	arg CGT	91n CAG	ala GCT	asp GAC	97.0	phe TTT	leu 17G	ئ م م
30	2 a E	ala CCT	\$ 5 5	ala CCT		tyr TAC	leu CTG	3 4
	p gly val GGT GTG	phe	asp glu CAT CAG	glu met GAA ATG	teu val arg	arg his pro AGA CAT CCT	177 978 1GC	
		818 CCC	asp	£ 8	√2. GTG	his CAT	a] a	phe TTT
	9-6 ACG	leu 11e TTG ATT	ala GCT	gly GGT	1eu 776	879 AGA	ala GCT	17.8 AA
05		leu 77G	val	tyr	£ 33	£ 5	2 X	g br
35		v.a.1 GTG	53 oys val TCT CTT (thr	5 2	a18 CCC	asp	CTC
•		21 8la leu val leu lle GCC TTG GTG TTG ATT	th ACA	95 SA	CTC CTC	fle ATT	ala GCT	ser leu gln lys phe gly AGT CTC CAA AAA TTT GGA
		21 818 GCC	2	81 årg CGT	111 88n AAC	181 8 8 8	171 818 GCT	201 ala GCC

5	(890)	(980)	(1070)	(1150)	(1250)	(1340)	(1430)	(1520)	(1610)	(002
	260 1eu CTT	290 11e ATT	320 ala CCT	350 ala GCC	380 leu CCT	\$10 810 CGT (440 his cat (470 ser AGT (*		530 val GTT (1700)
-	335	289 cys TGC		3 5	מני כ	val GTT O				
	a1a SC S	his CAC	asn tyr AAC TAT	47. A9.		leu v TTA G	638 cys lys TGT AAA	pro val CCA GTA	il pro T CCC	ala leu GCA CTT
10	asb arg GAC AGG	36. 77.	lys AAA	leu CTG	phe 1ys TTT AAA	leu 1 CTG T	437 4 cys c 1GT T		r val	ain thr ala leu CAA ACT GCA CTT
10	sso GAC	178 A	316 cys lys TCC AAA	5 5		ala J GCG C	∓3 ¥ ≥ ≥	s thr	thr tyr ACA TAC	aln thr
	\$ 7	£ ₹	val G17	leu leu CTG CTG	asp glu GAT GAA	asn ala AAT GCG	# 8 - 2	alu lys GAG AAA	alu thr tyr GAA ACA TAC	£.3
	5 F.	Tec	SAT	16.0			¥ ¥		o olu T CAA	
		Jeu	\$ \$	val val GTC GTG	val phe GTG TTC	phe aln TTC CAG	= 2 = 2	leu hís TTG CAT	1 asp C GAT	**
15	253 glu cys GAA TGT	pro leu leu CCT CTG TTG	ser lys asp AGT AAG GAT	ser val val leu leu arg TCT GTC GTG CTG CTG CTG AGA		lys p	₹ 5 ₹ 5	val le	glu val GAA GTC	oln ile iys iys CAA ATC AAG AAA
		17 17 17	of S	ty 1	ala lys val phe GCC AAA GTG TTC	F 2	£₹ ≿≴	55	2 9 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	£ 3
	250 1eu CTG	280 54 4 50 64 4 50	310 val glu ser lys asp val cys lys asm tyr GTT CAA AGT AAG GAT GTT TGC AAA AAC TAT	340 asp tyr GAT TAC	370 tyr 8 TAT G	400 gly glu tyr lys phe GCA GAG TAC AAA TTC	436 asn leu qly lys val qly ser lys cys AAC CTA GGA AAA GTG GGC AGC AAA TGT	460 461 gln leu cys val leu his CAG TTA TGT GTG TTG CAT	490 ala leu GCT CTG	520 ser glu lys glu arq gln lle lys lys TCT GAG AAG AGA CAA ATC AAG AAA
	asp GAT	279 cys TCT	phe	340 pro asp CCT GAT	369 370 cys tyr a TGC TAT (4 6 9 5 7 9	# # D	2 2 2	490 r ala A GCT	520 5 9 tu
		278 978 100	889 CAT	34 F	alu c		# ₹ 25	تا 19ء د 10ء	e ser T TCA	glu lys GAG AAG
20		£ \$	ala GCT	£ 55	his alu CAT GAA	gln leu CAG CTT	ser arg TCA AGA	Jeu asn CTG AAC	s phe C TTF	28
	246 678 160	1ys	818 CCT	ΕŞ	מנים הניד כ			= 5 = 5	pro cys phe CCA TCC TTT	2 Se
		leu CTG	ser leu ala ala TCA TTA GCT GCT	ale arg arg his pro GCA AGA AGG CAT CCT	asp p	phe alu TTT CAC	glu val GAG GTC	ser val val leu asn TCC GTG GTC CTG AAC		514 cys thr leu ser TGC ACA CTT TCT
	245 glu cys GAA TGC	lys leu AAA CTG	16. 16.	tyr a	e ele ccA			r val C 676	asn arg arg AAC AGG CGA	Ę Ç
25			670	glu 6	ala a cct	glu leu GAG CTT	leu val CTT GTA	u ser A TCC	n arg c AGG	
	hís thr CAC ACG	Ber Ber TCC AGT	leu pro TTG CCT	tyr glu tyr TAT GAA TAT	8 18 8 CC C	392 cys glu leu TGT GAG CTT	thr In	tyr leu TAT CTA	1 asn G AAC	asp 11e GAT ATA
		A to	8 8 CAC	1ec 170	361 cys a TGT		pro ti		480 ser leu val TCC TTG GTG	
~ . '	240 lys val AAA GTC	270 8er 11e TCG ATC	300 818 CCT	330 phe 1 TTT 1	360 3 0ys c TGC T	390 gln asn CAA AAT	420 thr pi ACT CC	450 glu asp GAA GAC	0 r leu c 11G	510 hís ala CAT GCA
		esp GAT		let 1	45 %		ser ti	450 8 91u A GAA	480 u ser A TCC	
30	leu thr CTT ACC	gin sap CAA GAT	met pro	gly met GGC ATG	76 A	1 5 7 A	# £		r gle	phe TTC
			914	leu TTG	thr leu glu lys ACT CTA GAG AAG	asn leu 11e lys AAT TTA ATC AAA	pro gln val ccc cAA GTG	pro oys ala ccc TGT GCA C	476 477 Cys cys thr TGC TGC ACA	glu thr phe thr GAA ACA TTC ACC
	thr ACA	265 ays glu esn TGT GAA AAT	252	- 2 5	7. J	. ŭ i	. 5 N	t pro	476 477 cys cys TGC TGC	1 phe 170
	val thr	265 0ys TGT	KT X	val phe GTC TTC	. F C	gla a CAG A		g met A ATG	47 10%	thr ACA
	77 Tec	15 N	23	asp v GAT G	glu thr GAA ACC	pro gin CCT CAG	lys val	A ACA	¥	glu
35	lys leu val thr asp AAG TTA GTG ACA GAT	tyr ile Tat atc	2 5 T	lys asp val phe leu AAG GAT GTC TTC TTG		glu pr	γ. Α.	A AAA	thr Acc	ala CCT
			291 ala glu val glu asn GCC GAA GTG GAA AAT	5 5 5 4		glu glu GAA GAG	thr lys lys val ACC AAG AAA GTA	glu ala lys arg GAA GCA AAA AGA	arg val thr lys AGA GTC ACC AAA	501 glu phe asn ala GAG TTT AAT GCT
	231 val ser GTT TCC	261 ala lys GCC AAG	291 818 9 GC G	321 glu ala GAG GCA	351 lys thr AAG ACA	381 val glu GTG GAA		g glu	ACA	phe
•	., , ,	·• • •	~ * 0	M D G	M#3	381 val GTG	411 tyr 1AC	949 970 CCT	471 889 CAC	91° 91°

5	(1790)	leu ter TTA TAA CATCACATTTAAAAG (1883)	AAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTTTT
	560 1ys	Ϋ́	¥
	558 559 560 0ys cys lys TGC TGC AAG	ter TAA	ICTT
	558 378	ACAŢ	AATT
	lys AAG	CATC	CATA
10	970	T &	AAA
	CTA C	TT A	CTAA
	phe TTT	753	CTGT
	ala GCT	leu TTA	CACC
15	als CCT	နှင့် ၁၂၁	CCAA
	TTC	ala GCT	AAAG
	asp GAT	580 914 CAA	CTCT
	550 1ys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys oys cys lys AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG AAG TGC TGC AAG	570 phe ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter TTT GCC GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA	CTTC
	met ATG	۾ 25	1110
20	val	ata CCT	ITCTI
	ela GCT	val GTT	arra
	1ys AA	Jeu CTT	VTCT
	1eu CTG	lys AA	VTTC/
25	gla	<u>₹</u> ¥) (11)
25	glu	41 <i>y</i>	WW.
	lys	g)u GGC	3ATC
	540 thr	570 910 646	Ğ
	818 GCA	a1 a GCC	¥¥
30	1 ys		X
·	510	567 973 100	9
	1ys AAG	thr	ter
	h1s CAC	95 68	ter VTGAGAA
	lys AAA	1ys AG	CCA1
35	leu val lys his lys pro CTC GTG AAA CAC AAG CCC	asp lys CAT AAG	XCTA
	leu CTC	esp GAC	ter ter Catctcagggtaggagataagaa
	531 91 ^u 6A6	561 818 GCT	27.0

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAAAAATGGAAAGAATCTAA.... 20AA (2078)

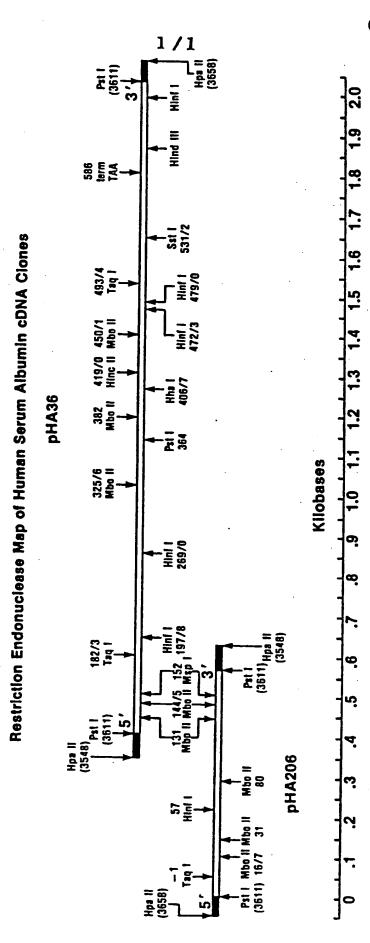
9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

5	(30)	(170)	(1260)	(350)	(660)	(330)	(620)	(017)	(300)
	AG Se	20 173 AA	50 818 974	80 CTT	110 Pro CCA	140	170 91n CAA	200 cys TCT (230 910 GAA (
	4 F	phe TTC	gye TT	thr ACT	AAC	1 E	169 cys TGC		8 ts 0
	1ec CTC	asn	₹ 3	# CC A CC	9.50 CAC	177	168 cys cys 151	leu lys CTC AAG	
10	phe	ale GAA	thr Act	val	asb CAT		- 2 d		alu phe GAG TTT
	3 5	. € ₹	CTA C	thr ACA	1ys	1ys	At A	ala a	ata q
	2 E	\$ 5	3	£ \$ 5	2 2		age 1	lys ain ara AAA CAG AGA	lys a
	10 T	Jeu TTG	AAT	3 5	d Ja	Dhe leu TTT TTG	ala GCT	818 GCC /	pro lys CCC AAA
	11e	asp GAT	val GTG	1ys AA	11G		ala u	36F 2	
15	p r o val tlu phe GTA ACC TTT	₹	3 ¥	S S S S S S S S S S S S S S S S S S S	phe TTC	glu thr GAG ACA	lys AAA	36. 700	arg phe AGA TTT
	val tlu GTA ACC	age 11			10 10 10 10 10	91°	tyr TAT		gln a
	2 5 E	10 his arg CAT CGG		ope TT	충흥중	130 88n AAT	160 ACC ACC	190 lys ala AAC GCT	220 867 9
	t 5			1e	AAT	SAC GAC		91. 555	75
20	1ye	SCT	88 CAT	thr	2 S	M.S.	ala lys CCT AAA	- 2 8	ala arg leu ccr ccc crc
20	Atc Atc	2. T2	3 8	CA T	919 GG	phe	ohe TT	esp GAT	818
		glu	ag E	CT CT	pro CCT	ala GCT	phe TTC	5 33	val GTA 0
		AGT	6 A	ser TCA	2 ₹	thr		leu	8 8 800
		lys AAG	4 8 E	₹ ₹	ele CAA	124 cys TGC	glu leu leu GAA CTC CTT	glu leu arg GAA CTT CGG	t 5
25		* 1 s	gln CAG	asp GAC	1 ys		g &	esp	61.8 A 23
		* 58	gla	62 cys TGT	*1*	va) GTG	5 20	leu CTC	
		CAT) eu	60 glu asn GAA AAT	91 693 TGT	asp	<u> </u>	Jys AG	phe lys TTC AAA
		- : 3	30 tyr 1AT		8 8 5 5	120 val GTT	150 tyr TAT	180 73 CCA	210 ala ccT
		phe erg	gln	ala GCT	asp GAC	e e e	phe TTT	1 c 1 c 1	<u>5</u> 3
30		-	ala GCT	Ser TCA	313 CCT	r Z	tyr TAC	Jeu CTG	28
		4 × 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	phe TTT	9 ¹⁶	glu met GAA ATG	8.7 ACA	pro CCT	177 8ys 1GC	
		-1 -6 p tyr ser srg gly vsl TAT TCC ACG GGT GTG	2 2 2 3 3 3 3 3	asp CAT	₹8	val arg GTG AGA	ala arg arg his pro GCC AGA AGA CAT CCT	177 Iys ale ale oys AAA GCT GCC TGC	phe TTT
		A 5 5 A	11e ATT	ala GCT	91y GGT	arg leu CGA TTG	5 2	als CCT	1ys AAA
35		- # E	35	val	tyr TAT	ř S	£ 3	13s	gin iys CAA AAA
- •		tyr TAT	21 ala leu val leu ile ala phe GCC TTG GTG TTG ATT GCC TT1	51 53 lys thr cys val ala asp glu AAA ACA TGT GTT GCT GAT GAG	thr	leu pro arg leu val arg CTC CCC CGA TTG GTG AGA		esp GAT	201 ala ser leu gin iya phe giy GCC AGT CTC CAA AAA TTT GGA
		ser ele TCG GCT	1ev 11G	51 lys thr AAA ACA	£ ₹	leu CTC	141 glu 11e GAA ATT	al a	AGT
		301	21 ala CCC	2	81 819 CGT	asn AAC	₹ ₹ ₹	171 818 GCT	201 818 GCC

	=	5	©	ć	Ę	ę	6	6	E	6
5	(B90)	(980)	(1070)	(1160)	(1250)	(13%0)	(1430)	(1520)	(1610)	(1700)
	260 Jeu CTT	290 11e ATT	320 ala G:T	350 ala ccc	380 teu CCT	410 818 CGT	0 4 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	470 ser AGT	500 175 AAA	530 val Gff
	939	289 cys 1GC	tyr TAT	3 5	oro CCT		\$ \ \	val GTA	٠ <u>٠</u> کان	leu CTT
	a la 800	h is	lys asn AAA AAC	87 70 70 70 70	1ys AAA	1es	438 cys TCT	970 CCA		878 GCA
10	50V		316 cys lys TGC AAA	leu arg CTG AGA	phe TTT	leu leu CTG TTA	437 cys TGT	thr Acc	tyr TAC	thr
•	esp asp arq ala GAT GAC AGG GCG	glu lys GAA AAA	316 val cys GTT TGC	Jeu	a to	asn ala AAT CCG	Tys AAA	\$ \$.	thr	£ &
	8SD GAT			leu CTG	asp CAT	phe gin asn ala leu leu val TTC CAG AAT CCG CTG TTA GTT	ser AGC	glu lys GAG AAA	5 5	1ys AAA
	8 t 8	leu leu CTG TTG	GAT	val	phe TTC	ala CAG	4) de 23	leu his TTG CAT	asp GAT	Jys AAG
	253 Cys TGT	leu CTG	ser lys asp AGT AAG CAT	val GTC	val	TTC	val GTG	1ec 77G	va! GTC	11e ATC
15	28	pro CCT	ser AGT	ser TCT	370 tyr ala lys TAT GCC AAA	1 xs	Jys AAA	va1 GTG	490 ala leu glu val asp qlu thr tyr val GCT CTG GAA GTC GAT GAA ACA TAC GTT	520 glu arq ain ile lys lys ain thr ala GAG AGA CAA ATC AAG AAA CAA ACT GCA
	250 leu leu CTG CTT	X X 3	916 AA	340 asp tyr GAT TAC	3 50 00 00 00 00 00 00 00 00 00 00 00 00	400 glu tyr GAG TAC	41y 66A	461 cys TGT	leu CTG	AGA AGA
	250 gly asp leu leu GGA GAT CTG CTT	280 CA 250 CA 250	310 phe val TTT GTT				420 thr pro thr leu val glu val ser arg asm leu gly lys val ACT CCA ACT CTT GTA GAG GTC TCA AGA AAC CTA GGA AAA GTG	\$60 gln leu CAG TTA	490 ala cct	520 91u GAG
	GAT O	278 279 1ys leu lys glu cys cys AAA CTG AAG GAA TGC TGT	phe TTT	his pro CAT CCT	369 qlu cys GAA TGC	alu aln leu aly GAG CAG CTT GGA	asn AAC		phe ser TTT TCA	leu ser glu lys g
20	his gly	278 cys	asp CAT		g g CAA	leu CTT	AGA AGA	tyr leu ser val val leu asn TAT CTA TCC GTG GTC CTG AAC		g)u GAG
20	3 5	lys glu AG GAA	818	arg arg AGA AGG	pro his CCT CAT	alu aln GAG CAG	ser	Jeu CTG	pro oys CCA TGC	ser TCT
	245 246 glu oys oys GAA TGC TGC	1ys	leu pro ser leu ala TIG CCT TCA TTA GCT	AGA			val GTC	val GTC	7 A	CTT CTT
	245 glu eys GAA TGC	lys leu AAA C7G	leu TTA	6. 6.	ala ala asp CCT GCA GAT	leu phe CTT TTT	910	val GTG	arg arg AGG CGA	Ş ţ
		. 178	15 T	glu tyr GAA TAT	# 55	leu CTT	val	36r 700	arg AGG	
25	his thr	ser AGT	pro	38	ala ala GCC GCT	g g g	leu CTT	leu CTA	AAC	11e ATA
		3er 700	1eu	tyr TAT		392 cys	thr		val GTG	asp GAT
	240 lys val AAA GTC	270 ser 11e TCG ATC	6 % CA %	1e0	360 361 oys cys TCC TCT	390 gln asn CAA AAT	5 A	450 glu esp GAA GAC	480 ser leu val TCC TTG GTG	510 hts ale asp tie CAT GCA GAT ATA
			300	330 Pre 111						
	leu thr CTT ACC	gin asp CAA GAT	pro	gly met GGC ATG	glu lys GAG AAG	1 ys	val ser GTG TCA	448 cys ala TGT GCA	thr glu ACA GAA	thr phe ACC TTC
30			ATG			. 11e	_			
	\$ 5	glu asn GAA AAT	916	16u	glu thr thr leu GAA ACC ACT CTA	leu TTA	at a	500	477 cys TGC	glu thr phe GAA ACA TTC
	ž V		asn asp AAT GAT	ahe TT	thr ACT	gin asn CAG AAT		ATG	476 0ys 100	glu thr GAA ACA
	510	265 9 998 1 TGT	e AT	val	ACC		val GTA	arg AGA	thr 1ys ACC AAA	
35	1ec	. 11e	val glu GTG GAA	asp CAT	95 8	5 5	lys lys AG AAA	1ys AAA	thr	asn ala AAT GCT
	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	265 lys tyr 11e gys AAG TAT ATC TGT	glu vel glu esn asp GAA GTG GAA AAT GAT	1ys	tyr TAT	glu glu pro GA GAG CCT	1ys AG	glu ala lys arg met pro GAA GCA AAA AGA ATG CCC	val	AAT
		AAG		321 glu ala lys asp val phe leu GAG GCA AAG GAT GTC TTG	JS1 lys thr tyr AAG ACA TAT		411 tyr thr lys lys val pro TAC ACC AAG AAA GTA CCC	88	8 4 5 8 4 5	501 glu phe GAG TTT
	231 val GTT	261 a1a GCC	291 918 000	321 91u 646	351 173 AAG	381 val GTG	411 tyr TAC	441 Pro	471 85P GAC	501 91u GAG

5	559 560 cys lys TCC AAG (1790)		(CTTTAA (2002)	
10	558 val glu lys cys c	leu ter ter TTA TAA CATCACATTTAAAAG (1883)	TCTAAAAACATAAATT	
15	550 asp phe ala ala phe val glu lys cys cys lys CAT TIC CCT CCT TIT GTA GAG AAG TCC TCC AAG	io n ala ala leu qiy A GCT GCC TTA GCC	GTAAAGCCAACACCCTG	
20	53 73 ala val met asp as 74 GCT GTT ATG GAT CA	570 glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter CAG CAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA	AAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTCTTTTTCGTTGGTGAAAGCCAACACCCTGTCTAAAAACATAAATTTCTTTAA (2002)	18/02) w 07 ·
25	540 lys ala thr lys glu gin leu lys ala val met asp a ANG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT o	glu qly lys lys le GAG GGT AAA AAA C1	AGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTCTTTTTCGTTGG	
30	540 ys pro lys ala thr AG CCC AAG GCA ACA	phe ala TTT GCC	er Aagagaagaaaatgaag Cottcaataaaaaa	
35	531 glu leu val lys his lys pro GAG CTC GTG AAA CAC AGG CCC	561 ala asp asp lys glu thr cys GCT GAC GAT AAG GAG ACC TGC	ter ter Catctcacctaccatgadataagaaa Tcattttgcctctttctctgtggttgaa	

- 10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
- 11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
 - 13. A DNA transfer vector according to claim 12, which is a plasmid.
- 14. A DNA transfer vector according to claim 13,10 wherein the plasmid is pBR322 or YEp6.
 - 15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
- 16. A DNA transfer vector according to any of15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
 - 17. A vector or process according to claim 16, wherein the bacterium or yeast is <u>E. coli</u> or <u>Saccharomyces cerevisiae</u>.



This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS
\square IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
☐ FADED TEXT OR DRAWING
☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
☐ SKEWED/SLANTED IMAGES
☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
☐ GRAY SCALE DOCUMENTS
☑ LINES OR MARKS ON ORIGINAL DOCUMENT
☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

IMAGES ARE BEST AVAILABLE COPY.

☐ OTHER:

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.